R E P O R T R E S U M E S

ED 020 084 RE 001 242

EMPIRICAL DETERMINATION OF THE INSTRUCTIONAL READING LEVEL.

BY- BORMUTH, JOHN R.

PUB DATE APR 68

EDRS PRICE MF-$0.25 HC-$0.32 6P.

DESCRIPTORS- *READABILITY, *CLOZE PROCEDURE, *READING COMPREHENSION, *READING MATERIAL SELECTION, INFORMAL READING INVENTORY, READING TESTS, REDUNDANCY, INDEPENDENT STUDY, INFORMATION THEORY, *INFORMATION PROCESSING, READING LEVEL, READING ASSIGNMENTS,

THE FIRST OF A SERIES OF STUDIES DESIGNED TO INVESTIGATE THE SUITABILITY OF INSTRUCTIONAL MATERIALS WITH REGARD TO LEVELS OF DIFFICULTY IS REPORTED. THE STUDY ATTEMPTED TO DETERMINE WHETHER SOME RANGE OF DIFFICULTY MAXIMIZED THE AMOUNT OF INFORMATION STUDENTS GAIN AS A CONSEQUENCE OF READING INSTRUCTIONAL MATERIALS. IN ORDER TO FORM 129 PAIRS ACCORDING TO READING ABILITY, STUDENTS RANGING FROM GRADE 3 THROUGH GRADUATE LEVEL WERE GIVEN A CLOZE TEST. ONE MEMBER OF EACH PAIR WAS GIVEN A CLOZE READABILITY TEST OVER A PASSAGE TO DETERMINE THE DIFFICULTY OF THE PASSAGE FOR THAT PAIR. THE OTHER MEMBER OF THE PAIR DETERMINED HOW MUCH INFORMATION THAT PAIR GAINED BY READING THE PASSAGE. HE FIRST GUESSED THE ANSWERS TO A MULTIPLE CHOICE TEST AND THEN READ THE PASSAGE AND TOOK THE SAME TEST AGAIN. INFORMATION GAIN WAS DETERMINED BY SUBTRACTING THE SCORE OF HIS FIRST TEST FROM THE SCORE OF THE SECOND. IT APPEARED FROM THE DATA OBTAINED THAT IT MAY BE POSSIBLE TO ESTABLISH FAIRLY DEFINITE STANDARDS OF WHAT IS A SUITABLE PASSAGE FOR USE BY A CHILD. IT WAS FOUND THAT SCORES ON CLOZE TESTS DO NOT DEPEND SOLELY ON A SUBJECT'S PRIOR KNOWLEDGE OF THE CONTENT OF A PASSAGE. EIGHT REFERENCES ARE GIVEN. THIS PAPER WAS PRESENTED AT THE INTERNATIONAL READING ASSOCIATION CONFERENCE (BOSTON, APRIL 24-27, 1968). (KJ)

IRA, BOSTON
APRIL 1968

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY John R. Bormuth
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

ED020084

Empirical Determination of the Instructional Reading Level*

J. R. Bormuth

Teachers are often admonished to give the child instruction only from materials which are suitable for him. Thorndike (7) first suggested standards of suitability which have come to be adopted in roughly this form. Materials are said to be suitable for use in the child's independent, unsupervised study if he can correctly answer at least 90 per cent of the comprehension questions asked him about the materials after he has studied them. The materials are regarded as suitable for use in supervised instruction if he can answer 75 per cent or more of the questions. When the child can answer fewer than 75 per cent of the questions, study of the materials is thought to result in frustrations which cause the child to learn negative attitudes toward instruction. Since a number of well known authors (see Betts (1), Bond and Tinker (2), and Harris (5)) advocate the use of these standards, presumably the standards are widely used by teachers. The problem is that neither Thorndike nor anyone known to the author has given either logical or empirical reasons why these standards should be accepted.

This problem is potentially too serious to be left unattended. To illustrate, materials might be difficult just because they contain information which is new to a reader. Barring a child from reading difficult materials might actually bar him from studying the very materials from which he would learn the most, thus creating boredom by forcing him to read dull, repetitious materials. Conversely, until we have definite evidence to the contrary, we cannot be certain that even materials on which children can answer 90 per cent of the questions are easy enough to prevent children from becoming frustrated and learning the negative attitudes which cause them to dislike and reject study.

This paper reports the first of a series of studies designed to investigate this problem. The present study asks simply if there is some range of difficulty which maximizes the amount of information students gain as a consequence of reading instructional materials. Other studies the author currently has under way are investigating the difficulty levels at which student interest is maximized.

Procedure

The general pattern of this study involved these operations. First, students were given a cloze test in order to form pairs of students matched in reading ability. Second, one member of each pair was given a cloze readability test over a passage to determine the difficulty of the passage for that pair of students. Third, the other member of the pair was used to determine how much information that pair gained by reading the passage. He first tried to guess the answers for a multiple choice test made from the passage and then later read the passage and took the same test again. Information gain was determined by subtracting his first score from the score he made the second time he took the test. The information gained by a pair of students was then plotted against the cloze difficulty of that passage for that pair of students to determine the shape of the curve.

RE001 242

* The author wishes to acknowledge the long hours of labor Mr. Paul Berger of the University of Minnesota contributed to constructing the materials and gathering the data for this study.

COPYRIGHT 1968

The cloze test used to match the pairs of students was made from a 263 word passage taken from an elementary psychology textbook by Cretch and Crutchfield (6). A 52 item cloze test was made from this passage by deleting every fifth word from the passage and replacing the deleted words with underlined blanks of 15 typewriter spaces. The students were told how the test was made and that their task was to figure out what word was left out of each blank and to write their response in the blank. Responses were scored correct only when they exactly matched the words deleted, ignoring misspellings which did not otherwise make the response ambiguous. The students were then ranked from high to low on the basis of their scores and then divided into pairs taking each successive pair of students and randomly assignir a member to each of the two groups designated hereafter as groups Y and Z.

A cloze readability test and a multiple choice comprehension test was made for eac of two passages designated passages A and B. Passage A contained 469 words and passage B contained 393 words. Each passage described in non-technical language a psychological experiment. Five cloze readability test forms were made for each passage by deleting words 1, 6, 11, etc., to make the first form, words 2, 7, 12, etc., to make the the second, and so on until all of the five forms possible had been made. These forms were later randomly assigned to subjects, so that all five forms were used equally often. The multiple choice tests made from passage A contained 34 items and the one made from passage B contained 39. Each item contained four alternative responses. The comprehension tests underwent three editorial revisions each time trying them out on small groups of students. When these tests were scored, the scores were corrected for guessing by subtracting one-third the number of incorrect responses from the number correct. The reliabilities of the cloze readability tests, pooling all five forms in a single split-half correlation, were .92 and .89 for passages A and B, respectively. The reliabilities of the comprehension tests for passages A and B were .34 and .36, respectively. Both sets of reliabilities were calculated using the data reported in this study, the scores from the second administration of the comprehension test being the ones used in that case. All reliability correlations were corrected for test length.

This study was designed to partially replicate its findings and to make maximum use of student testing time by using two different sets of passages and tests but the same set of students. Table 1 shows the order in which the tests were administered. The test sessions were always at least a week apart with none spaced more

Table 1
Design of Test Administration Schedule

Group	Test Session		
	1	2	3
X	Matching Test	Comprehension A and Cloze B	Read A and Comprehension A
Y	Matching Test	Comprehension B and Cloze A	Read B and Comprehension B

two weeks apart. No time limit was imposed on any of the tests. Using this testing design, the cloze score of member X of each pair was used to determine the cloze difficulty of passage B for his pair while the two comprehension scores of member Y were used to determine the information that pair gained as a consequence of reading passage B. Conversely, the cloze score of member Y was used to determine the cloze difficulty of passage A for that pair and the comprehension scores of member X were used to determine the information the pair gained from reading passage A.

Originally it was planned to use just second year junior college students in the study. However, it proved necessary to use students ranging from grade 3 through graduate level in order to obtain a sufficient range of cloze difficulty scores. A total of 130 pairs of students were tested, 25 pairs in grade 3, 23 in grade 5, 15 in grade 7, 20 in grade 11, 24 in junior college, and 15 pairs enrolled in a graduate course. Because of absences, the data presented for passage A are based on 129 pairs and those for passage B on 125 pairs.

Analysis and Results

All scores were converted to percentage scores and then the cloze difficulty score for each pair was correlated with the information gain score. Third degree polynomial curves were fit to the data yielding a multiple correlation .69 on passage and a correlation of .62 on passage B. The multiple correlations were calculated by determining the multiple correlations between the information gain scores and the cloze difficulty scores, their squares, and their cubes. Only the linear and quadratic terms contributed significant amounts of variance to these correlations.

In order to determine the shapes of the curves relating information gain to cloze difficulty, eighth degree polynomial curves were fit to both the data from passage A and passage B. These curves were highly similar having a correlation of .95 over the range of observed scores. In order to show the general shape of the relationship between information gain and cloze difficulty the two sets of data were combined and a single eighth degree polynomial curve fit to the data. This curve is shown in Figure 1. The curve shows that the pairs of students who made scores of less than 17 per cent on cloze readability tests on the passages also tended to exhibit little gain in information when they read the passages. But for pairs making cloze readability scores in the range between 17 and 37 per cent, there was a sharp increase in the information gained. As pairs made still higher cloze readability scores, their information gain scores tended to level off with, perhaps, come slight tendency to increase.

In order to assess the fit of this curve to the data, the scores were divided into 14 intervals on the cloze scale and the mean gain scores calculated at each interval. Table 2 shows these data. Note that nearly all of these means fall close to the curve. None of the differences between pairs of interval means from the two different passages was significant when t - tests were applied. Nor did one set of means tend to be consistently higher than the other set. While 8 of the 14 differences favored passage B, this result, when the sign test was applied, was not significant and there appeared to be no consistent tendency for the differences between the curves to concentrate just in one segement of the curve. Hence, the two different sets of materials seemed to yield roughly identical results. Where the means showed fairly large differences, the number of subjects on which the means were based tended to be small suggesting that the differences were due to this source of instability.

Table 2
Mean of the Information Gain Scores in Each Interval of the Cloze Difficulty Scores

Cloze Difficulty Interval	Mean Gain Score			
	Passage A		Passage B	
	Number	Mean	Number	Mean
0 - 4	7	.022	3	.024
5 - 9	9	.123	4	.121
10 - 14	10	.082	9	.149
15 - 19	10	.005	9	.100
20 - 24	8	.107	5	.178
25 - 29	3	.243	19	.218
30 - 34	13	.304	11	.389
35 - 39	5	.181	14	.442
40 - 44	5	.474	9	.429
45 - 49	15	.407	10	.451
50 - 54	16	.422	20	.439
55 - 59	12	.426	6	.519
60 - 64	12	.442	4	.354
65 - 69	4	.472	2	.470

Discussion

It appears from these data that it may be possible to establish fairly definite standards of what is or is not a passage suitable for use by a child. These data seem to show that a student can gain very little information from studying materials on which his cloze readability score is below 37 per cent and that using materials much easier than the 37 per cent level will permit the child to acquire only slightly more knowledge while reading them. However, these results should be regarded as preliminary. Only two passages were used and both of those were quite similar in difficulty. Further, it is not certain how well these results will generalize to comprehension tests made by other test writers.

The question of whether the traditional 75 and 90 per cent levels mentioned at the beginning of this report are adequate can be only partially answered from these data. Bormuth (3 and 4) found that the cloze readability test scores comparable to these two levels were 44 and 57 per cent, respectively. These levels are shown on Figure 1 by the broken vertical lines. It can be seen from this figure that students who are given instruction from materials at one of these levels gain little more information than students given instruction from materials at the other level. And both levels seem to result in near maximum information gain.

However, the question of whether the 44 and 57 per cent cloze readability criteria are desirable cannot be resolved without finding out how the difficulty of materials influences students' interests and attitudes toward the materials. It may be that the 44 or 75 per cent criterion is too stringent and that materials having cloze or comprehension test readabilities as low as 37 or 65 per cent, respectively, may be perfectly suitable for students when both interest and information gain are considered. Similarly, there is presently no way to be sure that

passages on which a student scores above the cloze and comprehension test readability criterion scores of 57 and 90 per cent, respectively, are not so simple to the student that he rejects them as being insipid and an insult to his intelligence. Finally, there is no reason to think that the criterion scores will necessarily be the same for students of all ages, on passages at all levels of difficulty, and on passages on all topics. Obviously this topic calls for a great deal more research.

On Cloze Text Theory

These data are also relevant to the question of whether cloze tests measure the redundancy of the passages from which they are made -- an important issue in cloze test theory. Weaver and Kingston (8) for example assert that cloze tests are largely a measure of redundancy but fail to make clear in just what sense they are using the term redundancy. In its most rigorous sense redundancy refers to the frequencies with which patterns of letters, words, parts of speech and so on occur in the language. It is a logical absurdity to assert that cloze tests measure the frequencies of occurrence of anything except the subject's responses. What they may have meant was that cloze tests measure either the subject's prior knowledge of the information contained in a passage or the familiarity of the language patterns of the subject. The present study shows rather decisively that scores on cloze tests do not depend solely upon a subject's prior knowledge of the content of a passage. As cloze scores increased, information gain increased. Further, scores on the comprehension test administered before the subjects had read the passage had correlations of .09 and .11 with cloze scores. These correlations were not significantly different from zero at the .05 level. If cloze scores do reflect the subject's prior knowledge of the information in the passages, the effect seemed so weak as to be negligible in the present study.

References

1. Betts, E. A. Foundations of Reading Instruction. New York: American Book, 1946.
2. Bond, G. L., and Tinker, M. A. Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts, 1967.
3. Bormuth, J. R. "Comparable Cloze and Multiple Choice Comprehension Test Scores," Journal of Reading, 10 (1967), 291-299.
4. Bormuth, J. R. "Cloze Test Readability: Criterion Reference Scores," Journal of Educational Measurement, (in press).
5. Harris, A. J. Effective Teaching of Reading. New York: David MacKay, 1962.
6. Kretch, D., and Crutchfield, R. S. Elements of Psychology. New York: Knopf, 1958.
7. Thorndike, E. L. "Reading and Reasoning: A Study of Mistakes in Paragraph Reading." Journal of Educational Psychology, 8 (1917), 323-332.
8. Weaver, W. W., and Kingston, A. J. "A Factor Analysis of the Cloze Procedure and Other Measures of Reading and Language Ability," Journal of Communication, 13 (1963), 252-261.

FIGURE 1

EIGHT DEGREE POLYNORMAL CURVE FITTED TO THE REGRESSION OF EACH PAIR'S INFORMATION GAIN SCORE ON ITS CLOZE READABILITY SCORE